Collegio Carlo Alberto

A class of neutral to the right priors induced by superposition of beta processes

Pierpaolo De Blasi
Stefano Favaro
Pietro Muliere

No. 130
December 2009

Carlo Alberto Notebooks

www.carloalberto.org/working_papers

A class of neutral to the right priors induced by superposition of beta processes

P. De Blasi[1], S. Favaro[2], P. Muliere[3]

[1] Università degli Studi di Torino and Collegio Carlo Alberto, Torino, Italy.
E-mail: pierpaolo.deblasi@unito.it

[2] Università degli Studi di Torino and Collegio Carlo Alberto, Torino, Italy.
E-mail: stefano.favaro@unito.it

[3] Università L. Bocconi, Milano, Italy.
E-mail: pietro.muliere@unibocconi.it

October 2009

Abstract

A random distribution function on the positive real line which belongs to the class of neutral to the right priors is defined. It corresponds to the superposition of independent beta processes at the cumulative hazard level. The definition is constructive and starts with a discrete time process with random probability masses obtained from suitably defined products of independent beta random variables. The continuous time version is derived as the corresponding infinitesimal weak limit and is described in terms of completely random measures. It takes the interpretation of the survival distribution resulting from independent competing failure times. We discuss prior specification and illustrate posterior inference on a real data example.

Key words and phrases: Bayesian nonparametrics, beta process, beta-Stacy process, completely random measures, neutral to the right priors, survival analysis

1 Introduction

Let $\mathbb{F}_{\mathbb{R}_+}$ be the space of all cumulative distribution functions on the positive real line. In this paper we introduce a stochastic process $\{F_t, t \geq 0\}$ with trajectories in $\mathbb{F}_{\mathbb{R}_+}$ which belongs to the class of neutral to the right (NTR) priors. A random distribution function F on $\mathbb{R}_+$ is NTR if, for any $0 \leq t_1 < t_2 < \ldots < t_k < \infty$ and for any $k \geq 1$, the random variables (r.v.s)

$$F_{t_1},\ \frac{F_{t_2} - F_{t_1}}{1 - F_{t_1}},\ \ldots\ ,\frac{F_{t_k} - F_{t_{k-1}}}{1 - F_{t_{k-1}}} \tag{1}$$

are independent, see Doksum (1974). NTR priors share some remarkable theoretical properties, among which the most celebrated one is the conjugacy with respect to right-censored survival data. The form of the posterior distribution and its large sample properties are now well known (see, e.g., Ferguson and Phadia (1979), Kim and Lee (2001, 2004). An interesting extension of NTR priors has been recently introduced by James (2006) with the family of spatial NTR processes.

NTR priors can be represented as suitable transformations of completely random measures (CRMs), i.e. random measures that give rise to mutually independent random variables when evaluated on pairwise disjoint sets. Appendix A.1 provides a brief account of CRMs, as well as justification of the following statements. It is important to recall that F is NTR if and only if $F_t = 1 - \mathrm{e}^{-\tilde{\mu}((0,t])}$ for some CRM $\tilde{\mu}$ on $\mathscr{B}(\mathbb{R}_+)$ (Borel σ-algebra of $\mathbb{R}_+$) such that $\mathbb{P}[\lim_{t\to\infty} \tilde{\mu}((0,t]) = \infty] = 1$. The nonatomic part of $\tilde{\mu}$ (that is the part without fixed jumps) is characterized by its Lévy intensity ν, which is a nonatomic measure on $\mathbb{R}_+ \times \mathbb{R}_+$, so that the law of F is uniquely determined by ν and the density of the fixed jumps. The conjugacy property of NTR priors can be then expressed as follows: the posterior distribution of F, given (possibly) right-censored data, is described by a NTR process for a CRM $\tilde{\mu}^*$ with fixed jump points at uncensored observations. This result is of great importance for statistical inference; indeed, the posterior distribution, conditional on right-censored data, is still NTR and one can fully describe the associated CRM in terms of the updated Lévy intensity and the densities of the jumps at fixed points of discontinuity. Therefore, one can resort to the simulation algorithm suggested in Ferguson and Klass (1972) to sample the trajectories of the underlying CRM, thus obtaining approximate evaluations of posterior inferences.

The beta-Stacy process of Walker and Muliere (1997) is an important example of NTR prior. Its main properties are (i) a parametrization with a straightforward interpretation in terms of survival analysis that facilitates prior specification; (ii) a simple description of the posterior process in terms of the parametrization used in the prior. For later reference, we recall these two properties. As for (i), we adopt the parametrization in Walker and Muliere (1997, Definition 3) and we suppose, as is usual in applications, that the underlying CRM $\tilde{\mu}$ does not have fixed jump points in the prior. To this end, let α be a diffuse measure on $\mathscr{B}(\mathbb{R}_+)$ and $\beta : \mathbb{R}_+ \to \mathbb{R}_+$ a piecewise continuous and positive function such that $\int_0^t \beta(x)^{-1}\alpha(\mathrm{d}x) \to +\infty$ as $t \to +\infty$. A beta-Stacy process $\{F_t, t > 0\}$ with parameter (α, β) is NTR for a CRM $\tilde{\mu}$ whose Lévy intensity is defined by

$$\nu(\mathrm{d}s, \mathrm{d}x) = \frac{\mathrm{d}s}{1 - \mathrm{e}^{-s}}\, \mathrm{e}^{-s\beta(x)}\, \alpha(\mathrm{d}x). \tag{2}$$

In particular, $\mathbb{E}(F_t) = 1 - \exp\{-\int_0^t \beta(x)^{-1}\alpha(\mathrm{d}x)\}$, see equation (32) in Appendix A.1, suggesting that $H_0(t) = \int_0^t \beta(x)^{-1}\alpha(\mathrm{d}x)$ takes on the interpretation of the prior guess at the cumulative hazard rate of F. The role played by α and β is better explained when one considers the nonparametric distribution induced by the beta-Stacy process on the space of cumulative hazard functions, i.e. the stochastic process $\{H_t, t > 0\}$ defined as

$$H_t = H_t(F) = \int_0^t \frac{\mathrm{d}F_x}{1 - F_{x^-}}. \tag{3}$$

It can be shown that $\{H_t, t > 0\}$ is distributed as a beta process of Hjort (1990) (see the forthcoming Remark 3.1), so that $\mathbb{E}(H_t) = H_0(t)$ and $\mathrm{Var}(H_t) = \int_0^t [\beta(x) + 1]^{-1} \mathrm{d}H_0(x)$. Then, β plays the role of concentration parameter: a large β makes for tighter concentration around H_0. As for (ii), consider data $(T_1, \Delta_1) \ldots, (T_n, \Delta_n)$ arising from n lifetimes subject to right censoring, where T stands for the time observed and Δ is the censoring indicator ($\Delta = 1$ indicates an exact observation, $\Delta = 0$ a censored one). We adopt a point process formulation, which is standard in survival analysis, by defining $N(t) = \sum_{i \leq n} \mathbb{1}_{(T_i \leq t, \Delta_i = 1)}$ and $Y(t) = \sum_{i \leq n} \mathbb{1}_{(T_i \geq t)}$. Based on this notation, one can describe the posterior distribution of F as a beta-Stacy process, that is $F_t|\text{data} = 1 - \mathrm{e}^{-\tilde{\mu}*((0,t])}$ where $\tilde{\mu}^*$ is a CRM with fixed jumps defined by (2) with updated parameter $(\alpha, \beta + Y)$ and fixed jumps $\{V_k,\ k \geq 1\}$ at locations $\{t_k,\ k \geq 1\}$ such that $1 - \mathrm{e}^{-V_k} \sim \mathrm{beta}(N\{t_k\}, \beta(t_k) + Y(t_k) - N\{t_k\})$. Here, $\mathrm{beta}(a, b)$ denotes the beta distribution and $N\{t_k\} = N(t_k) - N(t_k^-)$ is the number of uncensored observations occurring at time t_k.

Our aim is to introduce a new class of NTR priors and to investigate its properties with respect to (i) and (ii). The definition is constructive and starts with a discrete time process which satisfies the independence condition in (1). Following the idea of Walker and Muliere (1997), we adopt a stick breaking construction: let $0 < t_1 < t_2 < \ldots$ be a countable sequence of time points indexed by $k = 1, 2, \ldots$ and define

$$F_{t_k} = \sum_{j=1}^{k} V_j \prod_{l=1}^{j-1} (1 - V_l) \tag{4}$$

for $V_1, V_2, \ldots$ a sequence of independent r.v.s with values in the unit interval. Each V_j is recovered from the product of independent beta distributed r.v.s so that the conditional probability of an event at time t_k given survival at time t_{k-1} is the result of a series of m independent Bernoulli experiments. In Section 2, we discuss properties and possible simplifications of the proposed parametrization, then we provide formulas for the finite dimensional distributions. The continuous time version of the process is derived through a passage to the limit which leads to the specification of the underlying CRM in terms of a Lévy intensity of the form (see the forthcoming Theorem 2.1)

$$\nu(\mathrm{d}s, \mathrm{d}x) = \frac{\mathrm{d}s}{1 - \mathrm{e}^{-s}} \sum_{i=1}^{m} \mathrm{e}^{-s\beta_i(x)} \alpha_i(\mathrm{d}x).$$

The beta-Stacy process can be recovered as a particular case either by setting $m = 1$ or by taking $\beta_i = \beta$ for any i. In Section 3, we provide discussion on the proposed NTR prior by studying the induced distribution on the space of cumulative hazard functions. One obtains that the corresponding random cumulative hazard is given by the superposition of m independent beta processes (see the forthcoming Proposition 3.1), which motivates the name *m-fold beta NTR process* we will give to the new prior. It also suggests that the prior beliefs can be specified reasoning in terms of survival times generated by independent competing failure times. In Section 4, we give a complete description of the posterior distribution given right-censored data and we detail a Ferguson-Klass type simulation algorithm for obtaining approximate evaluation of posterior quantities for a real data example. In Section 5 some concluding remarks and future research lines are presented.

2 The m-fold beta NTR process

2.1 Discrete time construction

For $m \geq 1$, let us consider m sequences of positive real numbers $(\alpha_{1,\bullet}, \beta_{1,\bullet}) := \{(\alpha_{1,k}, \beta_{1,k}), k \geq 1\}, \ldots, (\alpha_{m,\bullet}, \beta_{m,\bullet}) := \{(\alpha_{m,k}, \beta_{m,k}), k \geq 1\}$ and m independent sequences of r.v.s $Y_{1,\bullet} := \{Y_{1,k}, k \geq 1\}, \ldots, Y_{m,\bullet} := \{Y_{m,k}, k \geq 1\}$ such that $Y_{i,\bullet}$ is a sequence of independent r.v.s with $Y_{i,k} \sim \text{beta}(\alpha_{i,k}, \beta_{i,k})$. Define the sequence of r.v.s $\{X_k, k \geq 1\}$ via the following construction:

$$\begin{aligned} X_1 &\overset{\mathrm{d}}{=} 1 - \prod_{i=1}^{m}(1 - Y_{i,1}) \\ X_2|X_1 &\overset{\mathrm{d}}{=} (1 - X_1)\left(1 - \prod_{i=1}^{m}(1 - Y_{i,2})\right) \\ &\vdots \\ X_k|X_1, \ldots, X_{k-1} &\overset{\mathrm{d}}{=} (1 - F_{k-1})\left(1 - \prod_{i=1}^{m}(1 - Y_{i,k})\right) \end{aligned} \tag{5}$$

where

$$F_k := \sum_{j=1}^{k} X_j$$

with the proviso $X_1 := X_1|X_0$. By using Theorem 7 in Springer and Thompson (1970) it can be checked that the conditional distribution of $X_k|X_1, \ldots, X_{k-1}$ is absolutely continuous with respect to the Lebesgue measure with density given by

$$\begin{aligned} f_{X_k|X_1,\ldots,X_{k-1}}(x_k|x_1, \ldots, x_{k-1}) &= \frac{1}{1 - \sum_{j=1}^{k-1} x_j} \prod_{i=1}^{m} \frac{\Gamma(\alpha_{i,k} + \beta_{i,k})}{\Gamma(\beta_{i,k})} \\ &\times G_{n,0}^{n,0}\left(\frac{1 - x_k}{1 - \sum_{j=1}^{k-1} x_j} \middle| \begin{array}{c} \alpha_{1,k} + \beta_{1,k} - 1, \ldots, \alpha_{m,k} + \beta_{m,k} - 1 \\ \beta_{1,k} - 1, \ldots, \beta_{m,k} - 1 \end{array} \right) \mathbb{1}_{(0<x_k<1)}. \end{aligned}$$

Here $G^{l,m}_{p,q}$ stands for the Meijer G-function. Refer to Erdélyi et al. (1953, Section 5) for a thorough discussion of the Meijer G-functions which are very general functions whose special cases cover most of the mathematical functions such as the trigonometric functions, Bessel functions and generalized hypergeometric functions. Under the construction (5), $X_k < 1 - F_{k-1}$ almost surely (a.s.), so that $F_k \leq 1$ a.s.. Moreover, we have

$$\mathbb{E}[F_k] = \left\{ \prod_{i=1}^{m} \frac{\alpha_{i,k}}{\alpha_{i,k} + \beta_{i,k}} \right\} + \left\{ \prod_{i=1}^{m} \frac{\alpha_{i,k}}{\beta_{i,k} + \beta_{i,k}} \right\} \mathbb{E}[F_{k-1}]$$

Based on this recursive relation, one can prove that a sufficient condition for $F_k \to 1$ a.s. is that $\prod_{k\geq 1} \prod_{i=1}^{m} \beta_{i,k}/(\alpha_{i,k} + \beta_{i,k}) = 0$. Hence, we can state the following result.

Lemma 2.1 *Let $\{t_k, k \geq 0\}$ be a sequence of time points in $\mathbb{R}_+$ with $t_0 := 0$ and let $\{F_t, t \geq 0\}$ be defined by $F_t := \sum_{t_k \leq t} X_k$ for any $t \geq 0$ according to construction* (5)*. If $F_0 = 0$ and*

$$\prod_{k\geq 1} \prod_{i=1}^{m} \left(1 - \frac{\alpha_{i,k}}{\alpha_{i,k} + \beta_{i,k}} \right) = 0,$$

then the sample paths of $\{F_t, t \geq 0\}$ belong to $\mathcal{F}_{\mathbb{R}_+}$ a.s.

Note that the random process $\{F_t, t \geq 0\}$ in Lemma 2.1 is a discrete time NTR random probability measure, see (1). We term $\{F_t, t \geq 0\}$ a *discrete time m-fold beta NTR*, according to the following definition.

Definition 2.1 *Let $\{X_k, k \geq 1\}$ be a sequence of r.v.s defined via construction* (5) *and let $\{t_k, k \geq 0\}$ be a sequence of time points in $\mathbb{R}_+$ with $t_0 := 0$. The random process $\{F_t, t \geq 0\}$ defined by $F_t := \sum_{t_k \leq t} X_k$ and satisfying conditions of Lemma 2.1 is a discrete time m-fold beta NTR process with parameter $(\alpha_{1,\bullet}, \beta_{1,\bullet}), \ldots, (\alpha_{m,\bullet}, \beta_{m,\bullet})$ and jumps at $\{t_k, k \geq 0\}$.*

Definition 2.1 includes as particular case the discrete time version of the beta-Stacy process. In fact, construction (5) is similar to the construction proposed in Walker and Muliere (1997, Section 3) which has, for any $k \geq 1$, $X_k | X_1, \ldots, X_{k-1} \stackrel{\mathrm{d}}{=} (1 - F_{k-1}) Y_k$ for $Y_k \sim \mathrm{beta}(\alpha_k, \beta_k)$. Hence, (5) generalizes the construction in Walker and Muliere (1997) by nesting for any $k \geq 1$ the product of independent beta distributed r.v.s.: the latter can be recovered by setting $m = 1$. Moreover, using some known properties for the product of independent beta distributed r.v.s, further relations between the two constructions can be established. We focus on a result that will be useful in the sequel and that can be proved by using a well known property of the product of beta r.v.s, see Theorem 2 in Jambunathan (1954).

Proposition 2.1 *A discrete time m-fold beta NTR process with parameter $(\alpha_{1,\bullet}, \beta_{\bullet}), (\alpha_{2,\bullet}, \beta_{\bullet} + \alpha_{1,\bullet}), \ldots, (\alpha_{m,\bullet}, \beta_{\bullet} + \sum_{i=1}^{m-1} \alpha_{i,\bullet})$ is a discrete time beta-Stacy process with parameter $(\sum_{i=1}^{m} \alpha_{i,\bullet}, \beta_{\bullet})$.*

The interpretation is as follows. The random quantity $X_k/(1-F_{k-1})$ represents the conditional probability of observing the event at time t_k given survival up to t_k. By construction (5), $X_k/(1-F_{k-1})$ is the result of m independent Bernoulli experiments: we observe the event if at least one of the m experiment has given a positive result, where the probability of success in the i-th experiment is $Y_{i,k} \sim \text{beta}(\alpha_{i,k}, \beta_{i,k})$. The particular parameter configuration $\beta_{i,k} = \beta_k + \sum_{j=1}^{i-1} \alpha_{j,k}$, $2 \leq i \leq m$, yields that the probability of at least one success is beta distributed, hence we recover the construction in Walker and Muliere (1997).

Let $\Delta^{(s)}$ denote the s-dimensional simplex, $\Delta^{(s)} = \{(x_1, \ldots, x_s) \in \mathbb{R}_+^s : \sum_{j=1}^s x_j \leq 1\}$. By the construction (5) and by using the solution of integral equation of type B in Wilks (1932), it can be checked that, for any integer s, the r.v.s $X_1, \ldots, X_s$ have joint distribution on $\Delta^{(s)}$ which is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^s$ with density given by

$$f_{X_1,\ldots,X_s}(x_1,\ldots,x_s) \propto \prod_{j=1}^{s} \left\{ \frac{x_j^{\bar{\alpha}_{1,j}-1}(1-\sum_{l=1}^{j} x_l)^{\beta_{m,j}-1}}{(1-\sum_{l=1}^{j-1} x_l)^{\bar{\alpha}_{1,j}+\beta_{m,j}-1}} \right.$$
$$\left. \times \int_{(0,1)^{m-1}} \prod_{i=1}^{m-1} v_i^{\alpha_{i,j}} (1-v_i)^{\bar{\alpha}_{i+1,j}-1} \left(1 - \frac{x_j[1-\prod_{l=1}^{i}(1-v_l)]}{1-\sum_{l=1}^{j-1} x_l}\right)^{c_{i,j}} \mathrm{d}v_i \right\} \mathbb{1}_{(x_1,\ldots,x_s)\in\Delta^{(s)}} \quad (6)$$

where

$$\bar{\alpha}_{i,j} := \sum_{l=i}^{m} \alpha_{l,j}, \quad c_{i,j} := \beta_{i,j} - (\beta_{i+1,j} + \alpha_{i+1,j}).$$

In particular, from (6) it can be checked that, for any $k \geq 1$, the r.v.s $X_1, X_2/(1-F_1), \ldots, X_k/(1-F_{k-1})$ are independent and $X_k/(1-F_{k-1}) \overset{\mathrm{d}}{=} 1 - \prod_{i=1}^{m}(1-Y_{i,k})$. Due to the more elaborated definition of the discrete time m-fold beta NTR process, the joint density (6) appears less manageable than in the case of the discrete time beta-Stacy process, i.e. the generalized Dirichlet distribution introduced in Connor and Mosimann (1969). However, in (6) one can recognize the generalized Dirichlet distribution multiplied by the product of integrals which disappears when $m = 1$ or under the condition of Proposition 2.1.

2.2 Infinitesimal weak limit

The next theorem proves the existence of the continuous version of the process as infinitesimal weak limit of a sequence of discrete time m-fold beta NTR processes. We start by considering the case of no fixed points of discontinuity.

Theorem 2.1 *Let $\alpha_1, \ldots, \alpha_m$, $m \geq 1$, be a collection of diffuse measures on $\mathscr{B}(\mathbb{R}_+)$ and let $\beta_1, \ldots, \beta_m$ be piecewise continuous and positive functions defined on $\mathbb{R}_+$ such that $\int_0^t \sum_{i=1}^m \beta_i(x)^{-1} \alpha_i(\mathrm{d}x) \to +\infty$ as $t \to +\infty$ for any i. Then, there exists a CRM $\tilde{\mu}$ without fixed jump points and Lévy intensity*

$$\nu(\mathrm{d}s, \mathrm{d}x) = \frac{\mathrm{d}s}{1-\mathrm{e}^{-s}} \sum_{i=1}^{m} \mathrm{e}^{-s\beta_i(x)} \alpha_i(\mathrm{d}x). \quad (7)$$

In particular, there exists a NTR process $\{F_t, t>0\}$ *defined by* $F_t = 1-\mathrm{e}^{-\tilde\mu((0,t])}$ *such that, at the infinitesimal level,* $\mathrm{d}F_t|F_t \overset{d}{=} (1-F_t)[1-\prod_{i=1}^m(1-Y_{i,t})]$ *where* $Y_{1,t},\ldots,Y_{m,t}$ *are independent r.v.s with* $Y_{i,t}\sim \mathrm{beta}(\alpha_i(\mathrm{d}t),\beta_i(t))$*.*

A detailed proof of Theorem 2.1 is deferred to Appendix A.2. The strategy of the proof consists in defining, for any integer n, the process $F_t^{(n)} = \sum_{k/n\le t} X_k^{(n)}$ where $\{X_k^{(n)}, k\ge 1\}$ is a sequence of r.v.s as in (5) upon the definition of m sequences $(\alpha_{1,\bullet}^{(n)},\beta_{1,\bullet}^{(n)}),\ldots,(\alpha_{m,\bullet}^{(n)},\beta_{m,\bullet}^{(n)})$ that suitably discretize $\alpha_1,\ldots,\alpha_m$ and $\beta_1,\ldots,\beta_m$ over the time grid $0,1/n,2/n,\ldots,k/n,\ldots$. By writing $F_t^{(n)}$ as $F_t^{(n)} = 1-\exp^{-Z_t^{(n)}}$ for $\{Z_t^{(n)}, t\ge 0\}$ the independent increments process defined by $Z_t^{(n)} = -\sum_{k/n\le t}\log[1-X_k^{(n)}/(1-F_{(k-1)/t}^{(n)})]$, the following limit as $n\to+\infty$ can be derived:

$$\mathbb{E}[\mathrm{e}^{-\phi Z_t^{(n)}}] \to \exp\left\{-\int_0^{+\infty}(1-\mathrm{e}^{-\phi s})\sum_{i=1}^m\int_0^t \mathrm{e}^{-s\beta_i(x)}\alpha_i(\mathrm{d}x)\frac{\mathrm{d}s}{1-\mathrm{e}^{-s}}\right\}, \tag{8}$$

which ensures the convergence of the finite dimensional distributions of $\{Z_t^{(n)}, t\ge 0\}$ to those of $\{\tilde\mu((0,t]), t\ge 0\}$ for a CRM with Lévy intensity in (7).

When the measures α_i have point masses, the limiting process is described in terms of a CRM $\tilde\mu$ with fixed jump points. Let $\{t_k, k\ge 1\}$ be now the sequence obtained by collecting all t_k such that $\alpha_i\{t_k\}>0$ for some $i=1,\ldots,m$ and let $\alpha_{i,c}$ be the non-atomic part of α_i. Then the limit in (8) becomes

$$\begin{aligned}\mathbb{E}[\mathrm{e}^{-\phi Z_t^{(n)}}] \to \exp\Bigg\{&-\int_0^{+\infty}(1-\mathrm{e}^{-\phi s})\sum_{i=1}^m\int_0^t \mathrm{e}^{-s\beta_i(x)}\alpha_{i,c}(\mathrm{d}x)\frac{\mathrm{d}s}{1-\mathrm{e}^{-s}}\\ &+\sum_{t_k\le t}\sum_{i=1}^m\int_0^{+\infty}(\mathrm{e}^{-\phi s}-1)\frac{\mathrm{e}^{-\beta_i(t_k)s}(1-\mathrm{e}^{-\alpha_i\{t_k\}s})}{s(1-\mathrm{e}^{-s})}\mathrm{d}s\Bigg\}\end{aligned}$$

where the second integral in the right hand side corresponds to $\log(\mathbb{E}[\mathrm{e}^{\phi\log(1-Y_{i,t_k})}])$ with $Y_{i,t_k}\sim \mathrm{beta}(\alpha_i\{t_k\},\beta_i(t_k))$, see Lemma 1 in Ferguson (1974). This motivates the following definition of a continuous time NTR process.

Definition 2.2 *Let* $\alpha_1,\ldots,\alpha_m$*,* $m\ge 1$*, be a collection of measures on* $\mathscr{B}(\mathbb{R}_+)$ *and let* $\beta_1,\ldots,\beta_m$ *be positive and piecewise continuous functions defined on* $\mathbb{R}_+$ *such that,*

$$\lim_{t\to+\infty}\int_0^t\frac{\alpha_i(\mathrm{d}x)}{\beta_i(x)+\alpha_i\{x\}} = +\infty, \quad i=1,\ldots,m. \tag{9}$$

The random process $\{F_t, t>0\}$ *is a* m*-fold beta NTR process on* $\mathbb{R}_+$ *with parameters* $(\alpha_1,\beta_1),\ldots,(\alpha_m,\beta_m)$ *if, for all* $t>0$*,* $F_t = 1-\mathrm{e}^{-\tilde\mu((0,t])}$ *for* $\tilde\mu$ *a CRM characterized by the Lévy intensity*

$$\nu(\mathrm{d}s,\mathrm{d}x) = \frac{\mathrm{d}s}{1-\mathrm{e}^{-s}}\sum_{i=1}^m \mathrm{e}^{-s\beta_i(x)}\alpha_{i,c}(\mathrm{d}x) \tag{10}$$

and fixed jump V_k at any t_k with $\alpha_i\{t_k\} > 0$ for some $i = 1, \dots, m$ so that V_k distributed according to

$$1 - \mathrm{e}^{-V_k} \stackrel{d}{=} 1 - \prod_{i=1}^{m}(1 - Y_{i,t_k}), \qquad Y_{i,t_k} \sim \mathrm{beta}(\alpha_i\{t_k\}, \beta_i(t_k)). \tag{11}$$

Using equation (32) in Appendix A.1, the prior mean of the survival function is recovered as

$$\begin{aligned}\mathbb{E}[1 - F_t] &= \exp\left\{ - \int_0^t \sum_{i=1}^{m} \frac{\alpha_{i,c}(\mathrm{d}x)}{\beta_i(x)} \right\} \prod_{t_k \le t} \prod_{i=1}^{m} \left(1 - \frac{\alpha_i\{t_k\}}{\beta_i(t_k) + \alpha_i\{t_k\}}\right) \\ &= \prod_{i=1}^{m} \prod_{[0,t]} \left(1 - \frac{\alpha_i(\mathrm{d}x)}{\beta_i(x) + \alpha_i\{x\}}\right). \end{aligned} \tag{12}$$

Note that, in the second equality, $\prod_{[0,t]}$ stands for the product integral operator. Condition (9) implies that (12) goes to zero when t grows to infinity, see Lemma 2.1 for a comparison with the discrete time case. Actually (9) implies more, namely that each of the m factors in (12) vanishes for $t \to +\infty$. In particular, (9) is consistent with the interpretation of $\int_0^t [\beta_i(x) + \alpha_i\{x\}]^{-1} \alpha_i(\mathrm{d}x)$ as a proper cumulative hazard function for each i. We will come back to this point later in Section 3.

Remark 2.1 The beta-Stacy process is a special case of Definition 2.2. It is clearly recovered by setting $m = 1$, cfr. Walker and Muliere (1997, Definition 3). Moreover, a second possibility is if we set, for $m \ge 2$, $\beta_i(x) = \beta(x) + \sum_{j=1}^{i-1} \alpha_j\{x\}$ for some fixed function $\beta(\cdot)$, then

$$\nu(\mathrm{d}s, \mathrm{d}x) = \frac{\mathrm{d}s}{1 - \mathrm{e}^{-s}} \sum_{i=1}^{m} \mathrm{e}^{-s\beta(x)} \alpha_{i,c}(\mathrm{d}x) = \frac{\mathrm{d}s}{1 - \mathrm{e}^{-s}}\ \mathrm{e}^{-s\beta(x)} \Big(\textstyle\sum_{i=1}^{m} \alpha_{i,c} \Big)(\mathrm{d}x),$$

and, for any t_k such that $\alpha_i\{t_k\} > 0$ for some $i = 1, \dots, m$, we have that the jump at t_k is distributed according to

$$1 - \mathrm{e}^{-V_k} \stackrel{\mathrm{d}}{=} 1 - \prod_{i=1}^{m}(1 - Y_{i,t_k}) \sim \mathrm{beta}\big(\textstyle\sum_{i=1}^{m} \alpha_i\{t_k\}, \beta(t_k) \big),$$

see Proposition 2.1. Hence, $\{F_t, t \ge 0\}$ is a beta-Stacy process with parameters $(\sum_{i=1}^{m} \alpha_i, \beta)$.

3 Superposition of beta processes

3.1 Prior on the space of cumulative hazards

In order to investigate further the properties of the m-fold beta NTR process, it is convenient to reason in terms of the induced prior distribution on the space of cumulative hazard functions.

In the sequel we rely on the key result that the random cumulative hazard generated by a NTR process can be described in terms of a CRM with Lévy intensity whose jump part is concentrated on $[0, 1]$, see Appendix A.1.

The most relevant example of nonparametric prior on the space of cumulative hazard functions is the beta process. According to Hjort (1990), a beta process $\{H_t, t > 0\}$ is defined by two parameters, a piecewise continuous function $c : \mathbb{R}_+ \to \mathbb{R}_+$ and a baseline cumulative hazard H_0 such that, if H_0 is continuous, $H_t = \tilde{\eta}((0, t])$ for a CMR $\tilde{\eta}$ without fixed jump points and Lévy intensity

$$\nu(\mathrm{d}v, \mathrm{d}x) = \mathbb{1}_{(0<v<1)} c(x) v^{-1} (1 - v)^{c(x)-1} \mathrm{d}v \, \mathrm{d}H_0(x). \tag{13}$$

The case of fixed points of discontinuity is accounted for by taking H_0 with jumps at $\{t_k, k \geq 1\}$ and $H_t = \tilde{\eta}((0, t])$ for $\tilde{\eta} = \tilde{\eta}_c + \sum_{k \geq 1} J_k \delta_{t_k}$ where (a) the Lévy intensity of $\tilde{\eta}_c$ is given by (13) after substituting H_0 for $H_0(t) - \sum_{t_k \leq t} H_0\{t\}$; (b) the distribution of the jump J_k at t_k is defined according to $J_k \sim \mathrm{beta}(c(t_k) H_0\{t_k\}, c(t_k)(1 - H_0\{t_k\}))$. The formulas for the mean and the variance of H_t are as follows, see Hjort (1990, Section 3.3),

$$\mathbb{E}(H_t) = H_0(t), \quad \text{and} \quad \mathrm{Var}(H_t) = \int_0^t \frac{\mathrm{d}H_0(x)[1 - \mathrm{d}H_0(x)]}{c(x) + 1}. \tag{14}$$

Remark 3.1 If $\{F_t, t > 0\}$ is a beta-Stacy process of parameter (α, β) and Lévy intensity given in (2), then

$$\nu_H(\mathrm{d}v, \mathrm{d}x) = \mathbb{1}_{(0<v<1)} v^{-1} (1 - v)^{\beta(x)-1} \mathrm{d}v \, \alpha(\mathrm{d}x)$$

see (33) in Appendix A.1. It turns out that ν_H corresponds to the Lévy intensity of the beta process of parameter (c, H_0) where $c(x) = \beta(x)$ and $H_0(t) = \int_0^t \beta(x)^{-1} \alpha(\mathrm{d}x)$. By inspection of Definition 3 in Walker and Muliere (1997) one sees that the conversion formulas, when the parameter measure α has point masses, become

$$c(x) = \beta(x) + \alpha\{x\} \quad \text{and} \quad H_0(t) = \int_0^t \frac{\alpha(\mathrm{d}x)}{\beta(x) + \alpha\{x\}}.$$

Let now $\{F_t, t \geq 0\}$ be a m-fold beta NTR process with parameter $(\alpha_1, \beta_1), \ldots, (\alpha_m, \beta_m)$ and α_i's diffuse measures. Then, by using (33) in Appendix A.1,

$$\begin{aligned} \nu_H(\mathrm{d}v, \mathrm{d}x) &= \mathbb{1}_{(0<v<1)} \frac{\mathrm{d}v}{v(1 - v)} \sum_{i=1}^m (1 - v)^{\beta_i(x)} \alpha_i(\mathrm{d}x) \\ &= \sum_{i=1}^m \mathbb{1}_{(0<v<1)} \beta_i(x) v^{-1} (1 - v)^{\beta_i(x)-1} \mathrm{d}v \, \frac{\alpha_i(\mathrm{d}x)}{\beta_i(x)} \end{aligned} \tag{15}$$

that is the sum of m Lévy intensities of the type (13). It follows that $H(F)$ is the superposition of m beta processes, according to

$$F_t \stackrel{\mathrm{d}}{=} 1 - \prod_{[0,t]} \left\{ 1 - \sum_{i=1}^m \mathrm{d}H_{i,x} \right\} \tag{16}$$

where $\{H_{i,t}, t > 0\}$ is a beta processes of parameter $(c_i, H_{0,i})$ where $c_i(x) = \beta_i(x)$ and $H_{0,i}(t) = \int_0^t \beta_i(x)^{-1}\alpha_i(\mathrm{d}x)$. Note that F can be seen as the distribution function of the minimum of m independent failure times,

$$F_t = \mathbb{P}\{\min(X_1, \ldots, X_m) \leq t\}, \quad \mathbb{P}(X_i \leq t) = 1 - \prod\nolimits_{[0,t]} \big(1 - \mathrm{d}H_{i,x}\big) \tag{17}$$

and $H_{i,x}$ takes the interpretation of the random cumulative hazard associated to the i-th failure type (*i-th failure-specific cumulative hazard*).

It is also interesting to see the similarity of (16) to the waiting time distribution in state 0 of a continuous time Markov chain $\{X_t, t > 0\}$ in the state space $\{0, 1, \ldots, m\}$ where 0 is the initial state and $H_{i,x}$ is the cumulative intensity of the transition from 0 to i, $i = 1, \ldots, m$, cfr. Andersen et al. (1993, Section II.7). Then $\mathbb{P}(X_t = 0) = \prod_{[0,t]}\{1 - \sum_{i=1}^m \mathrm{d}H_{i,x}\}$. The cumulative transition intensities are constrained to $\sum_{i=1}^m \mathrm{d}H_{i,x} \leq 1$ since, conditionally on the past, the transition out of state 0 in an infinitesimal time interval is the result of a multinomial experiment. However, in (16) the transition is rather the result of a series of independent Bernoulli experiments, which is equivalent to considering a competing risks model generated by independent latent lifetimes, see Andersen et al. (1993, Section III.1.2). The difference between the two representations is clarified when one consider the case of fixed points of discontinuity. By inspection of Definition 2.2, one has that (15) holds for $\alpha_{i,c}$ substituted for α_i and, for any t_k such that $\alpha_i\{t_k\} > 0$, $i = 1, \ldots, m$,

$$J_k := H_{t_k}(F) - H_{t_k^-}(F) = 1 - \mathrm{e}^{-V_k} \overset{\mathrm{d}}{=} 1 - \prod_{i=1}^m (1 - Y_{i,t_k}) \tag{18}$$

where $Y_{i,t_k} \sim \mathrm{beta}\big(\alpha_i\{t_k\}, \beta_i(t_k)\big)$ takes on the interpretation of the conditional probability $Y_{i,t_k} := P(X_i = t_k | X_i \geq t_k)$ according to (17). Hence J_k corresponds to the (random) probability that at least one success occurs in m independent Bernoulli trials with beta distributed probabilities of success. If the α_is have point masses in common, (18) can not be recovered from (16). In fact, instead of (18) we would have that $J_k \overset{\mathrm{d}}{=} \sum_{i=1}^m Y_{i,t_k}$ which is not in $[0, 1]$ unless exactly $m - 1$ of the beta jumps Y_{i,t_k} are identically zero. This suggests that, in general, (16) is not the correct way of extending the notion of superposition of independent beta processes at the cumulative hazard level since there is no guarantee that infinitesimally $\sum_{i=1}^m \mathrm{d}H_{i,t}$ takes values on the unit interval.

Remark 3.2 The condition that the beta processes $\{H_{i,t},\ t > 0\}$ and $\{H_{j,t},\ t > 0\}$ have disjoint sets of discontinuity points when $i \neq j$ implies that the jump J_k is beta distributed. Such an assumption is the device used in Hjort(1990, Section 5) for the definition of the waiting time distribution of a continuous time Markov chain with independent beta process priors for the cumulative transition intensities.

In order to derive the counterpart of (16) in the case of fixed points of discontinuity, we rewrite $\{H_{i,t},\, t \geq 0\}$ as $H_{i,t} = \tilde{\eta}_i((0,t])$ for a beta CRM $\tilde{\eta}_i$ defined as

$$\tilde{\eta}_i = \tilde{\eta}_{i,c} + \sum_{k \geq 1} Y_{i,t_k} \delta_{t_k} \tag{19}$$

where $\tilde{\eta}_{i,c}$ has Lévy intensity

$$\nu_i(\mathrm{d}v, \mathrm{d}x) = \mathbb{1}_{(0<v<1)} \beta_i(x) v^{-1} (1-v)^{\beta_i(x)-1} \mathrm{d}v \, \frac{\alpha_{i,c}(\mathrm{d}x)}{\beta_i(x)} \tag{20}$$

Then

$$\begin{aligned} F_t &\stackrel{\mathrm{d}}{=} 1 - \prod_{[0,t]} \left\{ 1 - \sum_{i=1}^{m} \tilde{\eta}_{i,c}(\mathrm{d}x) \right\} \times \prod_{t_k \leq t} \prod_{i=1}^{m} (1 - Y_{i,t_k}) \\ &= 1 - \prod_{i=1}^{m} \prod_{[0,t]} \{1 - \mathrm{d}H_{i,x}\} \end{aligned} \tag{21}$$

by writing the sum inside the product integral as a product (the CRMs $\tilde{\eta}_{i,c}$ cannot jump simultaneously). This is consistent with equation (12), hence with the interpretation of the m-fold beta NTR process $\{F_t, t > 0\}$ as the random distribution function of the minimum of m independent failure times, see (17). The following proposition clarifies how the m-fold beta NTR process corresponds to the superposition of beta processes at the cumulative hazard level in the presence of fixed points of discontinuity.

Proposition 3.1 *Let $\{F_t, t > 0\}$ be a m-fold beta NTR process with parameter $(\alpha_1, \beta_1), \ldots, (\alpha_m, \beta_m)$ and let $\{t_k, k \geq 1\}$ be the collection of time points such that $\alpha_i\{t_k\} > 0$ for some $i = 1, \ldots, m$. Also, let $\tilde{\eta}_1, \ldots, \tilde{\eta}_m$ be independent beta CRMs defined as in* (19)–(20)*. Then*

$$H_t(F) \stackrel{d}{=} \sum_{i=1}^{m} \tilde{\eta}_{i,c}((0,t]) + \sum_{k:\, t_k \leq t} \left(1 - \prod_{i=1}^{m} (1 - Y_{i,t_k}) \right).$$

Remark 3.3 Even the beta-Stacy process can be interpreted as a random distribution function of the minimum of m independent failure times. Actually, as a counterpart of Remark 2.1, the cumulative hazard of a beta-Stacy process of parameter (α, β) can be expressed as in Proposition 3.1 by decomposing the measure α as $\alpha(\mathrm{d}x) = \sum_{i=1}^{m} \alpha_i(\mathrm{d}x)$ (both the absolutely continuous part and point masses) and by defining $\beta_i(x) = \beta(x) + \alpha_i\{x\}$. However, the m independent beta CRMs $\tilde{\eta}_i$ are constrained to have similar concentration around the corresponding prior means, cfr. equation (14), whereas the m-fold beta NTR does not suffer from such a restriction.

3.2 Prior specification

We exploit the description of F as the random distribution in presence of m independent competing risks, see (21), aiming at expressing different prior beliefs for the m different failure-specific lifetime distributions. We start by considering the case of no fixed points of discontinuity and we assume all α_i's to be absolutely continuous on $\mathscr{B}(\mathbb{R}_+)$. Suppose we model the random failure-specific cumulative hazards $\{H_{i,t}, t > 0\}$, $i = 1, \dots, m$, by specifying the prior guess of the i-th failure-specific cumulative hazard to be $H_{0,i}(t) = \int_0^t h_{0,i}(x)\mathrm{d}x$. For k_i a positive integer, the parameter choice

$$\alpha_i(\mathrm{d}t) = k_i h_{0,i}(t)\mathrm{e}^{-H_{0,i}(t)}\mathrm{d}t, \qquad \beta_i(t) = k_i\mathrm{e}^{-H_{0,i}(t)}. \tag{22}$$

gives to k_i a prior sample size interpretation: with independent and identically distributed (iid) survival times from $H_{0,i}$, $\beta_i(t)$ may be interpreted as the number at risk at t in an imagined prior sample of uncensored survival times, with k_i the sample size, see Hjort (1990, Remark 2B). Different k_i's allow to specify different degrees of prior beliefs on each of the m components $H_{0,i}$, see Remark 3.3, while keeping the prior mean of the cumulative hazard equal to the sum $\mathbb{E}[H_t(F)] = \sum_{i=1}^m H_{0,i}(x)$.

A different prior specification of the α_i's and β_i's parameters is possible by resorting to the methods set forth in Walker and Muliere (1997, Section 2.1), which consist in specifying the uncertainty of the random distribution function F_t about its center by assigning arbitrarily the second moment. Let $\mu(t) = -\log\{\mathbb{E}(1 - F_t)\}$ and $\lambda(t) = -\log\{\mathbb{E}[(1 - F_t)^2]\}$, both assumed to be derivable. Since the α_i's are absolutely continuous, $\mu(t)$ coincides with the prior guess of the cumulative hazard and $\mu'(t) = \sum_{i=1}^m h_{0,i}(t)$ where $h_{0,i}(t)\mathrm{d}t = \beta_i(t)^{-1}\alpha_i(\mathrm{d}t)$. The quantity $\lambda(t)$ can be also decomposed into a sum: by using the Lévy-Khinchine representation (31) in Appendix A.1, one finds that $\lambda(t) = \sum_{i=1}^m \lambda_i(t)$ for $\lambda_i(t) = \int_0^t\int_0^\infty(1 - \mathrm{e}^{-2s})\frac{\mathrm{e}^{-s\beta_i(x)}}{1-\mathrm{e}^{-s}}\mathrm{d}s\,\alpha_i(\mathrm{d}x)$. Note that λ_i corresponds to the second moment of the random distribution function of the i-th failure-specific lifetime. Then, for each i, α_i and β_i are defined in terms of $h_{0,i}$ and λ_i as follows

$$\beta_i(t) = \frac{\lambda_i'(t) - h_{0,i}(t)}{2h_{0,i}(t) - \lambda_i'(t)}, \qquad \alpha_i(\mathrm{d}t) = \beta_i(t)h_{0,i}(t)\mathrm{d}t.$$

It is interesting to consider the application in a meta analysis experiment where one specifies the prior on the random distribution function on the basis of former posterior inferences. In this context, a m-fold beta NTR process with fixed points of discontinuity will be needed. Consider a system of two components, where each of them is subject to independent failure. The system fails when the the first component experiences a failure, so that, denoting by X_1 and X_2 the failure times specific to component 1 and 2, respectively, the system lifetime is given by $T = \min(X_1, X_2)$. Suppose that estimation on the distribution of X_1 and X_2 have been performed on two initial samples by using a beta process prior in each case. For $i = 1, 2$, let the posterior distribution of the cumulative hazard of X_i be described by the updated parameters $(c_i + Y_i, H_i^*)$, where

$$H_i^*(t) = \int_0^t \frac{c_i(x)\mathrm{d}H_{0,i}(x) + \mathrm{d}N_i(\mathrm{d}x)}{c_i(x) + Y_i(x)}$$

In the equation above, N_i and Y_i refer to the point process formulation of the i-th initial sample, possibly including right-censored observations, while $H_{0,i}$ is the prior guess for the cumulative hazard of X_i. Suppose we are now given a new sample of failure times of the system where the type of component which has caused the failure is not specified. We can draw inference on the distribution F of T by specifying the prior according to a m-fold beta NTR process with parameters

$$\alpha_i(\mathrm{d}x) = c_i(x)\mathrm{d}H_{0,i}(x) + \mathrm{d}N_i(\mathrm{d}x), \qquad \beta_i(x) = c_i(x) + Y_i(x) - N_i\{x\}$$

for $i = 1, 2$. This corresponds to specify $F_t = 1 - \mathrm{e}^{-\tilde{\mu}((0,t])}$ for a CRM $\tilde{\mu}$ with fixed jump points characterized by the Lévy intensity

$$\nu(\mathrm{d}s, \mathrm{d}x) = \frac{\mathrm{d}s}{1 - \mathrm{e}^{-s}} \sum_{i=1}^{m} \mathrm{e}^{-s[c_i(x)+Y_i(x)]} c_i(x)\mathrm{d}H_{0,i}(x)$$

and jump V_k at time t_k distributed according to

$$1 - \mathrm{e}^{-V_k} \overset{\mathrm{d}}{=} 1 - \prod_{i=1}^{m}(1 - Y_{i,t_k}), \qquad Y_{i,t_k} \sim \mathrm{beta}\big(N_i\{t_k\}, c_i(t_k) + Y_i(t_k) - N_i\{t_k\}\big).$$

4 Illustration

4.1 Posterior distribution

We start with the derivation of the posterior distribution given a set of possibly right censored observations. Let $\{F_t, t \geq 0\}$ be a m-fold beta NTR process with parameters $(\alpha_1, \beta_1), \ldots, (\alpha_m, \beta_m)$ and α_i diffuse measure for any i. Consider right-censored data $(T_1, \Delta_1) \ldots, (T_n, \Delta_n)$ summarized by the point processes $N(t) = \sum_{i \leq n} \mathbb{1}_{(T_i \leq t, \Delta_i = 1)}$ and $Y(t) = \sum_{i \leq n} \mathbb{1}_{(T_i \geq t)}$. In view of Theorem 5.1 in Appendix A.1, the posterior distribution of F is given by a NTR process for a CRM with fixed jump points at exact observations, $\tilde{\mu}^* \overset{\mathrm{d}}{=} \tilde{\mu}_c^* + \sum_{k:N\{t_k\}>0} V_k^* \delta_{t_k}$ where $\tilde{\mu}_c^*$ has Lévy intensity

$$\nu^*(\mathrm{d}s, \mathrm{d}x) = \frac{\mathrm{d}s}{1 - \mathrm{e}^{-s}} \sum_{i=1}^{m} \mathrm{e}^{-s[\beta_i(x)+Y(x)]} \alpha_i(\mathrm{d}x), \tag{23}$$

(cfr. (34) in Appendix A.1) while the density f_{t_k} of the jump V_k^* at time point t_k such that $N\{t_k\} > 0$ is given by

$$f_{t_k}(s) = \kappa(1 - \mathrm{e}^{-s})^{N\{t_k\}-1} \sum_{i=1}^{m} \mathrm{e}^{-s[\beta_i(t_k)+Y(t_k)-N\{t_k\}]}, \tag{24}$$

where κ is the appropriate normalizing constant (cfr. (35) in Appendix A.1). Note that $\tilde{\mu}_c^*$ can be described as a m-fold beta NTR process with updated parameters $(\alpha_1, \beta_1+Y), \dots, (\alpha_m, \beta_m+Y)$. However, the densities f_{t_k} have not the form in (11). Upon definition of

$$\alpha_i^*(\mathrm{d}x) = \alpha_i(\mathrm{d}x) + \delta_{N\{x\}}(\mathrm{d}x), \quad \beta_i^*(x) = \beta_i(x) + Y(x) - N\{x\}, \quad i = 1, \dots, m,$$

the distribution of V_k^* can be expressed as a mixture of beta r.v.s,

$$1 - \mathrm{e}^{-V_k^*} | I = i \sim \mathrm{beta}(\alpha_i^*\{t_k\}, \beta_i^*(t_k)) \tag{25}$$

$$\mathbb{P}(I = i) = \frac{B(\alpha_i^*\{t_k\}, \beta_i^*(t_k))}{\sum_{j=1}^m B(\alpha_j^*\{t_k\}, \beta_j^*(t_k))}, \quad i = 1, \dots, m \tag{26}$$

where $B(a, b)$ is the beta function $B(a, b) = \int_0^1 v^{a-1}(1-v)^{b-1}\mathrm{d}v$. It can be checked that the posterior mean of F is given by

$$\mathbb{E}(F_t|\mathrm{data}) = 1 - \exp\left\{-\int_0^t \sum_{i=1}^m \frac{\alpha_i(\mathrm{d}x)}{\beta_i(x) + Y(x)}\right\} \times \prod_{t_k \le t}\left\{1 - \frac{\sum_{i=1}^m B(\alpha_i^*\{t_k\} + 1, \beta_i^*(t_k))}{\sum_{i=1}^m B(\alpha_i^*\{t_k\}, \beta_i^*(t_k))}\right\}, \tag{27}$$

which provides a Bayes estimator for F. The Kaplan-Meier estimator $\hat{F}(t) := 1 - \prod_{t_k \le t}\{1 - N\{t_k\}/Y(t_k)\}$ is obtained when the concentration of the prior becomes arbitrarily small. To see this, it is convenient to reason in terms of the variance of the failure-specific cumulative hazards $H_{i,t}$ going to zero for each i. This in turn requires that both β_i and α_i go to zero, cfr. formula (14), so that (27) reduces to $\hat{F}(t)$.

4.2 Simulation from the posterior

In this section we detail how to simulate a trajectory from the posterior NTR process $\{1 - \mathrm{e}^{-\tilde{\mu}^*((0,t])}, t > 0\}$ for $\tilde{\mu}^*$ the CRM defined in (23) and (24). In the literature, there are a few algorithms that can be used to generate sample paths of a NTR process via the corresponding CRM, see, e.g., Wolpert and Ickstadt (1998), Lee and Kim (2004) and Lee (2007). Here we use the *Ferguson and Klass* algorithm, see Walker and Damien (2000) for a discussion.

Let T be the largest value of t for which we are interested in simulating the process and $\tilde{\mu}^{*,T}$ the CRM $\tilde{\mu}^*$ restricted on the interval $[0, T]$. The jump V_k^* at a fixed points of discontinuity t_k can be generated according to the mixture of beta density defined in (25)-(26). As for the part of $\tilde{\mu}^{*,T}$ without fixed points of discontinuity, say $\tilde{\mu}_c^{*,T}$, following Ferguson and Klass (1972) it can be expressed as

$$\tilde{\mu}_c^{*,T} = \sum_{k \ge 1} J_k \delta_{\tau_k}$$

where the random jumps J_k can be simulated in a decreasing order according to their size. Specifically, let $\mathrm{d}\nu_t^*(s) = \nu^*(\mathrm{d}s, (0,t])$, then the J_k's are obtained as solution of the equation $\theta_k = M(J_k)$, where $M(s) = \nu_T^*([s, +\infty))$ and $\theta_1, \theta_2, \ldots$ are jump times of a standard Poisson process at unit rate. The random locations τ_k's are obtained according to the distribution function

$$\mathbb{P}(\tau_k \leq t | J_k) = n_t(J_k) \quad \text{where} \quad n_t(s) = \frac{\mathrm{d}\nu_t^*}{\mathrm{d}\nu_T^*}(s) \tag{28}$$

Hence, one can set τ_k as the solution of $u_k = n_{\tau_k}(J_k)$ where $u_1, u_2, \ldots$ are iid from the uniform distribution on the interval $(0,1)$, independent of the J_ks.

The implementation of the algorithm requires both the calculation of integrals (single in $\mathrm{d}\nu_t^*(s)$ and double in $M(s)$) and the solution of equations such as $\theta_k = M(J_k)$ and $u_k = n_{\tau_k}(J_k)$. Some devices can be used to avoid, in part, the recourse to numerical subroutines. The measure $\mathrm{d}\nu_t^*(s)$ involves a sum of integrals of the type $\int_0^t \mathrm{e}^{-s[\beta_i(x)+Y(x)]}\alpha_i(\mathrm{d}x)$ which have closed form for a prior specification such as in (22) (note that $Y(x)$ is a step function). Then, we can avoid to solve numerically the equation $u_k = n_{\tau_k}(J_k)$ if we write $n_t(J_k)$ (the cumulative distribution of $\tau_k | J_k$ when seen as a function of t, see (28)) in a mixture form as follows

$$n_t(s) = \sum_{i=1}^{m} \omega_i(s) n_{i,t}(s)$$

for weights $\omega_i(s) = \nu_i(\mathrm{d}s, (0,T]) / \sum_{j=1}^m \nu_j(\mathrm{d}s, (0,T])$, ν_i the Lévy intensity of a beta-Stacy process of parameter $(\beta_i + Y, \alpha_i)$, see (2), and $n_{i,t}(s) = \nu_i(\mathrm{d}s, (0,t])/\nu_i(\mathrm{d}s, (0,T])$. Then, conditionally on $I = i$, where $\mathbb{P}(I = i) = \omega_i(J_k)$, τ_k can be generated by solving the equation $u_k = n_{i,\tau_k}(J_k)$, which can be done analytically under the prior specifications (22). As for the computation of $M(s)$, note that it can be written as

$$M(s) = \sum_{i=1}^{m} \int_0^T B_{\mathrm{e}^{-s}}(\beta_i(x) + Y(x), 0)\alpha_i(\mathrm{d}x)$$

where $B_z(a,b) = \int_0^z s^{a-1}(1-s)^{b-1}\mathrm{d}s$ denotes the incomplete beta function. $B_z(a,0)$ can be computed as the limit of the rescaled tail probabilities of a beta r.v.,

$$B_z(a,0) \approx B(a,\epsilon) \cdot \mathbb{P}(Y \leq z), \quad Y \sim \mathrm{beta}(a,\epsilon) \tag{29}$$

for ϵ small. The second integration in $M(s)$ and the solution of the equation $\theta_k = M(J_k)$ needs to be done numerically.

4.3 Real data example

As an illustrative example, we consider the Kaplan and Meier (1958) data set, already extensively used by many authors in the Bayesian nonparametric literature. The data consists of the lifetimes 0.8, 1.0^*, 2.7^*, 3.1, 5.4, 7.0^*, 9.2, 12.1^*, where * denotes a right censored observation. The prior on the random distribution function F is specified by a m-fold beta NTR process with $m = 2$ and parameters as in (22) with $k_1 = k_2 = 1$ and $h_{0,i}(t) = \frac{\kappa_i}{\lambda_i}\left(\frac{t}{\lambda_i}\right)^{\kappa_i - 1} \mathrm{e}^{-(t/\lambda_i)^{\kappa_i}}$, $i = 1, 2$, that is two hazard rates of the Weibull type. We choose $\lambda_1 = \lambda_2 = 20$, $\kappa_1 = 1.5$ and $\kappa_2 = 0.5$, so that the prior process is centered on a survival distribution with non monotonic hazard rate, see Figure 1.

[Figure 1 about here]

We sample 5000 trajectories from the posterior process on the time interval $[0, T]$ for $T = 50$. In the implementation of the Ferguson and Klass algorithm we set $\epsilon = 10^{-15}$ in the approximation of $B_{\mathrm{e}^{-s}}(\beta_i(x) + Y(x), 0)$, see (29), and we truncate the number of jumps in $\tilde{\mu}_c^{*,T}$ by keeping only the jumps which induce a relative error in the computation of M smaller than 0.001. It results in a smallest jump of order $2\mathrm{e}^{-15}$. We evaluate the posterior distribution of the probability F_t for $t = 5$ and the exponential-type functional of $\tilde{\mu}^*$,

$$I_T(\tilde{\mu}^*) = \frac{1}{1 - \mathrm{e}^{-\tilde{\mu}^*((0,T])}} \left[\int_0^T \mathrm{e}^{-\tilde{\mu}^*((0,t])} \mathrm{d}t - T\mathrm{e}^{-\tilde{\mu}^*((0,T])} \right]$$

which corresponds to the random mean of the distribution obtained via normalization of F_t over $[0, T]$. $I_T(\tilde{\mu}^*)$ can be used to approximate, for T large, the random mean of the posterior NTR process which takes the interpretation of expected lifetimes. $T = 50$ can be considered sufficiently large since direct computation using formula (27) leads to $\mathbb{E}(F_T|\text{data}) = 0.996$. The reader is referred to Epifani, Lijoi and Prünster (2003) for a study on the distribution of the mean of a NTR distribution function where the same data set is used with a beta-Stacy process prior. Note that the Ferguson and Klass algorithm is not implementable for generating a trajectory of F_t on the entire positive real axis. In fact, $\nu_T^*([s, +\infty)) \to +\infty$ as $T \to +\infty$ for any s unless the β_i functions explode at infinity, which is unlikely to be adopted in applications where one typically takes decreasing β_i's in order to induce a decreasing concentration of the prior distribution for large time horizons.

Figure 2(a) displays the histogram and the kernel density estimate of the posterior distribution of F_t for $t = 5$ (sample mean 0.3041, sample standard deviation 0.1587), while in Figure 2(b) we compare the density of the distribution of $I_T(\tilde{\mu}^*)$ with the density of $I_T(\tilde{\mu})$, the latter calculated over 5000 trajectories of the prior process. Sample mean and standard deviation for $I_T(\tilde{\mu}^*)$ are 10.7619 and 4.1692, respectively, while $I_T(\tilde{\mu})$ has mean 8.9093 and standard deviation 6.2368.

[Figure 2 about here]

5 Concluding remarks

In the present paper we have introduced and investigated the properties of a new NTR prior, named m-fold beta NTR process, for the lifetime distribution which corresponds to the superposition of independent beta processes at the cumulative hazard level. The use of the proposed prior is justified in presence of independent competing risks, therefore, it finds a natural area of application in reliability problems, where such an assumption is often appropriate. The typical situation is a system consisting of m components that fail independently from each other. The lifetime of the system is determined by the first component failure, so that the cumulative hazard results into the sum of the m failure-specific cumulative hazard. The m-fold beta NTR process allows to specify different prior beliefs for the components' failure time distribution and represents a suitable extension of the beta-Stacy process to this setting.

An interesting development consists in studying the case of a system failing when at least k out of the m components experience a failure ($k > 1$). The lifetime T of the system would correspond to the k-th smallest value among the m component failure times $X_1, \ldots, X_m$. It is no more appropriate to model the distribution of T with a NTR process since the conditional probability of a failure time at t does depend on the past, namely on how many components have experience a failure up to time t. One can still put independent NTR priors on the distributions of the X_i's and study the induced nonparametric prior for the distribution of T. In the simple case of $X_1, \ldots, X_m$ iid with common random distribution F, the random distribution of T is described by the process $F_t^{(k)} := \sum_{j=k}^{m} \binom{m}{j} [F_t]^j [1 - F_t]^{m-j}$. Further work is needed to establish the existence of the corresponding nonparametric prior.

Future work will also focus on adapting the idea of superposition of stochastic processes at the cumulative hazard level into a regression framework. The goal is to provide a Bayesian nonparametric treatment of the additive regression model of Aalen (1989), which specifies the hazard rate of an individual with covariate vector $\mathbf{z} = (z_1, \ldots, z_p)$ as the sum $h(t, \mathbf{z}) = h_0(t) + \gamma_1(t)z_1 + \ldots + \gamma_p(t)z_p$, for h_0 a baseline hazard and γ_i's the regression functions. There are two main issues to address this task. First, the shapes of the regression functions γ_i's are left completely unspecified, therefore they are not constrained to define proper hazards. Secondly, the nonparametric prior on h_0 and γ_i can not be taken as independent because of the restriction imposed by $h(t, \mathbf{z}) \geq 0$. Work on this is ongoing.

Appendix

A.1 NTR priors and CRMs

Here we review some basic facts on the connections between NTR priors and CRMs. The reader is referred to Lijoi and Prünster (2008) for an exhaustive account. Denote by $\mathcal{M}$ the space of boundedly finite measures on $\mathscr{B}(\mathbb{R}_+)$ (that is μ in $\mathcal{M}$ has $\mu((0,t]) < \infty$ for any finite t) endowed with the Borel σ-algebra $\mathscr{M}$. A CRM $\tilde{\mu}$ on $\mathscr{B}(\mathbb{R}_+)$ is a measurable mapping from some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ into $(\mathcal{M}, \mathscr{M})$ and such that, for any collection of disjoint sets $A_1, \ldots, A_n$ in $\mathscr{B}(\mathbb{R}_+)$, the r.v.s $\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_n)$ are mutually independent. This entails that the random distribution function induced by $\tilde{\mu}$, namely $\{\tilde{\mu}((0,t]), t \geq 0\}$, is an independent increment process. CRMs are discrete measures with probability 1 as they can always be represented as the sum of two components:

$$\tilde{\mu} = \tilde{\mu}_c + \sum_{k=1}^{M} V_k \delta_{x_k} \tag{30}$$

where $\tilde{\mu}_c = \sum_{i \geq 1} J_i \delta_{X_i}$ is a CRM where both the positive jumps J_i's and the locations X_i's are random, and $\sum_{k=1}^{M} V_k \delta_{x_k}$ is a measure with random masses $V_1, \ldots, V_M$, independent from $\tilde{\mu}_c$, at fixed locations $x_1, \ldots, x_M$. Finally, $\tilde{\mu}_c$ is characterized by the *Lévy-Khintchine* representation which states that

$$\mathbb{E}\left[\mathrm{e}^{-\int_{\mathbb{R}_+} f(x)\tilde{\mu}_c(\mathrm{d}x)}\right] = \exp\left\{ -\int_{\mathbb{R}_+ \times \mathbb{R}_+} \left[1 - \mathrm{e}^{-sf(x)}\right] \nu(\mathrm{d}s, \mathrm{d}x) \right\} \tag{31}$$

where f is a real-valued function $\tilde{\mu}_c$-integrable almost surely and ν, referred to as the Lévy intensity of $\tilde{\mu}_c$, is a nonatomic measure on $\mathbb{R}_+ \times \mathbb{R}_+$ such that $\int_B \int_{\mathbb{R}_+} \min\{s, 1\} \nu(\mathrm{d}s, \mathrm{d}x) < \infty$ for any bounded B in $\mathscr{B}(\mathbb{R}_+)$.

As pointed out by Doksum (1974), a random distribution function F is NTR (see definition in (1)) if and only if

$$F_t = 1 - \mathrm{e}^{-\tilde{\mu}((0,t])}, \qquad t \geq 0$$

for some CRM $\tilde{\mu}$ on $\mathscr{B}(\mathbb{R}_+)$ such that $\mathbb{P}[\lim_{t \to \infty} \tilde{\mu}((0,t]) = \infty] = 1$. We will use the notation $F \sim \mathrm{NTR}(\tilde{\mu})$. A consequence of this characterization is that, by using (31), the expected value of F_t is expressed in terms of the Lévy intensity ν of $\tilde{\mu}$ (no fixed jumps case) as

$$\mathbb{E}[F_t] = 1 - \mathbb{E}\big[\mathrm{e}^{-\tilde{\mu}((0,t])}\big] = 1 - \exp\Big\{ -\int_0^t \int_0^\infty (1 - \mathrm{e}^{-s}) \nu(\mathrm{d}s, \mathrm{d}x) \Big\}. \tag{32}$$

A second characterization of NTR prior in terms of CRMs arise while assessing the prior distribution induced by F on the space of cumulative hazards, see Hjort (1990). Let $F \sim \mathrm{NTR}(\tilde{\mu})$ for a CRM $\tilde{\mu}$ without fixed jumps and let $\nu(\mathrm{d}s, \mathrm{d}x) = \nu(s, x)\mathrm{d}s\,\mathrm{d}x$ (with a little abuse of notation) be the corresponding Lévy intensity. The random cumulative hazard $H(F)$, see (3), is given by

$$H_t(F) = \int_0^t \frac{\mathrm{d}F_x}{1 - F_{x^-}} = \tilde{\eta}((0,t]), \qquad t \geq 0$$

where $\tilde{\eta}$ is a CRM with Lévy intensity $\nu_H(\mathrm{d}v, \mathrm{d}x) = \nu_H(v, x)\mathrm{d}v\,\mathrm{d}x$ such that $\nu_H(v, x) = 0$ for any $v > 1$. The conversion formula for deriving the Lévy intensity of $\tilde{\eta}$ from that of $\tilde{\mu}$ is as follows:

$$\nu_H(v, x) = \frac{1}{1-v}\,\nu(-\log(1-v), x), \quad (v, x) \in [0, 1] \times \mathbb{R}_+ \tag{33}$$

that is $\nu_H(\mathrm{d}v, \mathrm{d}x)$ is the distribution of $(s, x) \mapsto (1 - \mathrm{e}^{-s}, x)$ under ν, see Dey, Erickson and Ramamoorthi (2003).

Consider now an exchangeable sequence of lifetimes $(X_i)_{i\geq 1}$ such that the law of the sequence is directed by a NTR process F for some CRM $\tilde{\mu}$,

$$X_i|F \overset{\text{iid}}{\sim} F \quad i \geq 1 \qquad F \sim \mathrm{NTR}(\tilde{\mu})$$

We derive the posterior distribution of F given $X_1, \dots, X_n$ subject to censoring times $c_1, \dots, c_n$, which can be either random or non–random. The actual data consist in the observed lifetimes $T_i = \min(X_i, c_i)$ and the censoring indicators $\Delta_i = \mathbb{1}_{(X_i \leq c_i)}$. Define $N(t) = \sum_{i\leq n} \mathbb{1}_{(T_i\leq t, \Delta_i=1)}$ and $Y(t) = \sum_{i\leq n} \mathbb{1}_{(T_i\geq t)}$, where $N(t)$ counts the number of events occurred before time t and $Y(t)$ is equal to the number of individuals at risk at time t.

Theorem 5.1 (Ferguson and Phadia (1979)). *Suppose $F \sim NTR(\tilde{\mu})$ where $\tilde{\mu}$ has no fixed jump points. Then the posterior distribution of F, given $(T_1, \Delta_1), \dots, (T_n, \Delta_n)$ is $NTR(\tilde{\mu}^*)$ with*

$$\tilde{\mu}^* = \tilde{\mu}_c^* + \textstyle\sum_{k: N\{t_k\}>0} V_k^* \delta_{t_k}$$

where $\tilde{\mu}_c^*$ is a CRM without fixed jump points and it is independent of the jumps V_k^*'s, which occur at the exact observations.

Let $\nu(\mathrm{d}s, \mathrm{d}x) = \rho_x(s)\mathrm{d}s\,\alpha(\mathrm{d}x)$ be the Lévy intensity of $\tilde{\mu}$. Then the Lvey measure ν^* of μ_c^* is given by

$$\nu^*(\mathrm{d}s, \mathrm{d}x) = \mathrm{e}^{-sY(x)}\rho_x(s)\mathrm{d}s\,\alpha(\mathrm{d}x) \tag{34}$$

whereas the density of the jump V_k^* at time point t_k is given by

$$f_{t_k}(s) = \frac{(1-\mathrm{e}^{-s})^{N\{t_k\}}\mathrm{e}^{-s[Y(t_k)-N\{t_k\}]}\rho_{t_k}(s)}{\int_0^\infty (1-\mathrm{e}^{-u})^{N\{t_k\}}\mathrm{e}^{-u[Y(t_k)-N\{t_k\}]}\rho_{t_k}(u)\mathrm{d}u}. \tag{35}$$

A.2 Proof of Theorem 2.1.

Following the lines of the proof of Theorem 2 in Walker and Muliere (1997) we define, for any integer n, m sequences of positive real numbers $(\alpha_{1,\bullet}^{(n)}, \beta_{1,\bullet}^{(n)}), \dots, (\alpha_{m,\bullet}^{(n)}, \beta_{m,\bullet}^{(n)})$ such that

$$\alpha_{i,k}^{(n)} := \alpha_i\left(\frac{k-1}{n}, \frac{k}{n}\right], \qquad \beta_{i,k}^{(n)} := \beta_i\left(\frac{k}{n} - \frac{1}{2n}\right)$$

for $i = 1, \ldots, m$ and for any $k \geq 1$. Moreover, let us consider m independent sequences $Y^{(n)}_{1,\bullet}, \ldots, Y^{(n)}_{m,\bullet}$ such that $Y^{(n)}_{i,\bullet}$ is a sequence of independent r.v.s with $Y^{(n)}_{i,k} \sim \mathrm{beta}(\alpha^{(n)}_{i,k}, \beta^{(n)}_{i,k})$. Based on this setup of r.v.s, for any n we define the random process $Z^{(n)} := \{Z^{(n)}_t, t \geq 0\}$ as

$$Z^{(n)}_t := -\sum_{k/n \leq t} \log\left(1 - \frac{X^{(n)}_k}{1 - F^{(n)}_{k-1}}\right)$$

with $Z^{(n)}_0 := 0$ and $\{X^{(n)}_k, k \geq 1\}$ a sequence of r.v.s defined according to (5).

The first step consists in showing that the sequence of random processes $\{Z^{(n)}_t, t \geq 0\}_{n\geq 1}$ converges weakly to the process $\{\tilde{\mu}((0,t]), t \geq 0\}$ for $\tilde{\mu}$ the CRM having Lévy intensity in (7). Let $\Gamma(x) = \int_0^\infty y^{x-1}\mathrm{e}^{-y}\mathrm{d}y$ be the gamma function. By using the recursive relation $\Gamma(x) = (x-1)\Gamma(x-1)$ and the Stirling formula $\Gamma(x) \cong (2\pi x)^{1/2}(x/e)^x$ when x is large, we have

$$\begin{aligned}
\log(\mathbb{E}[\mathrm{e}^{-\phi Z^{(n)}_t}]) &= \log\left(\mathbb{E}\left[\mathrm{e}^{-\phi \sum_{k/n\leq t}\sum_{i=1}^m -\log(1-Y^{(n)}_{i,k})}\right]\right) \\
&= \sum_{k/n\leq t}\sum_{i=1}^m \log \frac{\Gamma(\alpha^{(n)}_{i,k} + \beta^{(n)}_{i,k})\Gamma(\beta^{(n)}_{i,k} + \phi)}{\Gamma(\beta^{(n)}_{i,k})\Gamma(\alpha^{(n)}_{i,k} + \beta^{(n)}_{i,k} + \phi)} \\
&= \sum_{k/n\leq t}\sum_{i=1}^m \log \prod_{j=0}^{r-1} \frac{(\beta^{(n)}_{i,k} + j)(\alpha^{(n)}_{i,k} + \beta^{(n)}_{i,k} + \phi + j)}{(\alpha^{(n)}_{i,k} + \beta^{(n)}_{i,k} + j)(\beta^{(n)}_{i,k} + \phi + j)} \frac{\Gamma(\alpha^{(n)}_{i,k} + \beta^{(n)}_{i,k} + r)\Gamma(\beta^{(n)}_{i,k} + \phi + r)}{\Gamma(\beta^{(n)}_{i,k} + r)\Gamma(\alpha^{(n)}_{i,k} + \beta^{(n)}_{i,k} + \phi + r)} \\
&= \sum_{k/n\leq t}\sum_{i=1}^m \log \prod_{j\geq 0} \frac{(\beta^{(n)}_{i,k} + j)(\alpha^{(n)}_{i,k} + \beta^{(n)}_{i,k} + \phi + j)}{(\alpha^{(n)}_{i,k} + \beta^{(n)}_{i,k} + j)(\beta^{(n)}_{i,k} + \phi + j)} \\
&= \sum_{k/n\leq t}\sum_{i=1}^m \int_0^{+\infty} (\mathrm{e}^{-\phi s} - 1)\frac{\mathrm{e}^{-\beta^{(n)}_{i,k} s}(1 - \mathrm{e}^{-\alpha^{(n)}_{i,k} s})}{s(1 - \mathrm{e}^{-v})}\mathrm{d}s \\
&= \int_0^{+\infty} \frac{\mathrm{e}^{-\phi s} - 1}{s(1 - \mathrm{e}^{-s})} \sum_{k/n\leq t}\sum_{i=1}^m \mathrm{e}^{-\beta^{(n)}_{i,k} s}(1 - \mathrm{e}^{-\alpha^{(n)}_{i,k} s})\mathrm{d}s.
\end{aligned}$$

Since for $i = 1, \ldots, m$, as $n \to +\infty$,

$$\sum_{k/n\leq t} \mathrm{e}^{-\beta^{(n)}_{i,k} s}(1 - \mathrm{e}^{-\alpha^{(n)}_{i,k} s}) \to s \int_0^t \mathrm{e}^{-s\beta_i(x)}\alpha_i(\mathrm{d}x),$$

one has that, as $n \to +\infty$,

$$\log(\mathbb{E}[\mathrm{e}^{-\phi Z^{(n)}_t}]) \to \int_0^{+\infty} \frac{\mathrm{e}^{-\phi s} - 1}{1 - \mathrm{e}^{-s}} \sum_{i=1}^m \int_0^t \mathrm{e}^{-s\beta_i(x)}\alpha_i(\mathrm{d}x)\mathrm{d}s.$$

This result ensures the convergence of the finite dimensional distributions of $\{Z^{(n)}_t, t \geq 0\}$ to those of $\{\tilde{\mu}((0,t]), t \geq 0\}$, cfr. Lévy-Khintchine representation (31). The tightness of the sequence $(Z^{(n)})_{n\geq 1}$ follows by the same arguments used in Walker and Muliere (1997).

For any $n \in \mathbb{N}$, let us define the discrete time m-fold beta NTR process $\{F_t^{(n)}, t \geq 0\}$ such that $F_t^{(n)} := \sum_{k/n \leq t} X_k^{(n)}$ with $F_0^{(n)} = 0$. Since

$$-\log(1 - F_t^{(n)}) = -\sum_{k/n \leq t} \log\left(\prod_{i=1}^{m}(1 - Y_{i,k})\right) = -\sum_{k/n \leq t} \log\left(1 - \frac{X_k^{(n)}}{1 - F_{k-1}^{(n)}}\right) = Z_t^{(n)}$$

then $F_t^{(n)} = 1 - \mathrm{e}^{Z_t^{(n)}}$. Note that, for $t = k/n$, $F_{t^-}^{(n)} = F_{(k-1)/n}^{(n)}$ and $\mathrm{d}F_t^{(n)} := F_{t+\mathrm{d}t}^{(n)} - F_{t^-}^{(n)}$ is given by $F_{k/n}^{(n)} - F_{(k-1)/n}^{(n)}$ for $\mathrm{d}t$ small, where

$$F_{k/n}^{(n)} - F_{(k-1)/n}^{(n)} | F_{(k-1)/n}^{(n)} \overset{\mathrm{d}}{=} \left(1 - F_{(k-1)/n}^{(n)}\right)\left(1 - \prod_{i=1}^{m}(1 - Y_{i,k}^{(n)})\right).$$

The same arguments used when taking the limit above yields, at the infinitesimal level, $\mathrm{d}F_t | F_t \overset{\mathrm{d}}{=} (1 - F_{t^-})\,(1 - \prod_{i=1}^{m}(1 - Y_{i,t}))$ where $F_t = 1 - \mathrm{e}^{-\tilde{\mu}((0,t])}$ and $Y_{1,t}, \ldots, Y_{m,t}$ are independent r.v.s such that $Y_{i,t} \sim \mathrm{beta}(\alpha_i(\mathrm{d}t), \beta_i(t))$. The fact that $\{F_t, t \geq 0\}$ defines a random distribution function is assured by $\mathbb{E}[F_t] \to 0$ as $t \to +\infty$, which can be checked by using (32) under the hypothesis on $\int_0^t \sum_{i=1}^m \beta_i(x)^{-1}\alpha_i(\mathrm{d}x)$ since $\mathbb{E}[1 - F_t] = \exp\{-\int_0^t \sum_{i=1}^m \beta_i(x)^{-1}\alpha_i(\mathrm{d}x)\}$. It follows that $\{F_t, t \geq 0\}$ is a NTR process according to definition (1), which completes the proof. $\square$

Acknowledgements

The authors are grateful to Fabrizio Ruggieri for a useful discussion. This work is partially supported by MIUR grant 2008MK3AFZ.

References

[1] Aalen, O.O. (1989). A linear regression model for the analysis of life times. *Stat. Med.* **8**, 907–925.

[2] Andersen, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. (1993). *Statistical models based on counting processes.* Springer, New York.

[3] Connor, R. J. and Mosimann, I. E. (1969). Concepts of independence for proportions with a generalization of the Dirichlet distribution. J. Amer. Statist. Assoc. 64, 194-206.

[4] Dey, J., Erickson, R.V. and Ramamoorthi, R.V. (2003). Some aspects of neutral to the right priors. Internat. Statist. Rev. 71, 383–401.

[5] Doksum, K. A. (1974). Tailfree and neutral random probabilities and their posterior distributions. Ann. Probab. 2, 183–201.

[6] Epifani, I., Lijoi, A., Prünster, I. (2003). Exponential functionals and means of neutral-to-the-right priors. Biometrika 90, 791–808.

[7] Erdélyi, A., Magnus, W., Oberhettinger, F., and Tricomi, F. G. (1976). *Higher trascendental functions Vol. I.* McGraw-Hill, New York.

[8] Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2**, 615–629.

[9] Ferguson, T. S., Phadia, E. G. (1979). Bayesian nonparametric estimation based on censored data. Ann. Statist. 7, 163–186.

[10] Ferguson, T. S., Klass, M. J. (1972). A representation of independent increment processes without Gaussian components. Ann. Math. Statist. 43, 1634–1643.

[11] Hjort, N. L. (1990). Nonparametric Bayes estimators based on Beta processes in models for life history data. Ann. Statist. 18, 1259–1294.

[12] Jambunathan, M.V. (1954). Some properties of beta and gamma distributions. Ann. Math. Stat. 25, 401–405.

[13] James, L. F. (2006). Poisson calculus for spatial neutral to the right processes. Ann. Statist. 34, 416–440.

[14] Kaplan, E.L., Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Am. Statist. Assoc. 53, 457–481.

[15] Kim, Y., Lee, J. (2001). On posterior consistency of survival models. Ann. Statist. 29, 666–686.

[16] Kim, Y., Lee, J. (2004). A Bernstein-von Mises theorem in the nonparametric right-censoring model. Ann. Statist. 32, 1492–1512.

[17] Lee, J. (2007). Sampling methods for neutral to the right processes. J. Comput. Graph. Statist. 25, 401–405.

[18] Lee, J., Kim, Y. (2004). A new algorithm to generate beta processes. Comput. Statist. Data Anal. 25, 401–405.

[19] Lijoi, A., Prünster, I. (2008). Models beyond the Dirichlet process. In *Bayesian Nonparametrics* (Eds. N.L. Hjort, C.C. Holmes, P. Müller and S.G. Walker). Cambridge University Press, in press.

[20] Springer, M.D., Thompson, W.E. (1970). The distribution of products of Beta, Gamma and Gaussian random variables. SIAM J. Appl. Math. 18, 721–737.

[21] Walker, S.G. and Damien P. (2000). Representation of Lévy processes without Gaussian components. Biometrika 87, 477–483.

[22] Walker, S. G., Muliere, P. (1997). Beta-Stacy processes and a generalization of the Polya-urn scheme. Ann. Statist. 25, 1762–1780.

[23] Wilks, S.S. (1932). Certain generalizations in the analysis of variance. Biometrika. 24, 471–494.

[24] Wolpert, R. L., Ickstadt, K. (1998). Simulation of Lévy random fields. In *Practical Nonparametric and Semiparametric Bayesian Statistics* (Eds. D. Dey, P. Müller and D. Sinha), volume 133 of *Lecture Notes in Statistics*. Springer, New York, 227–242.

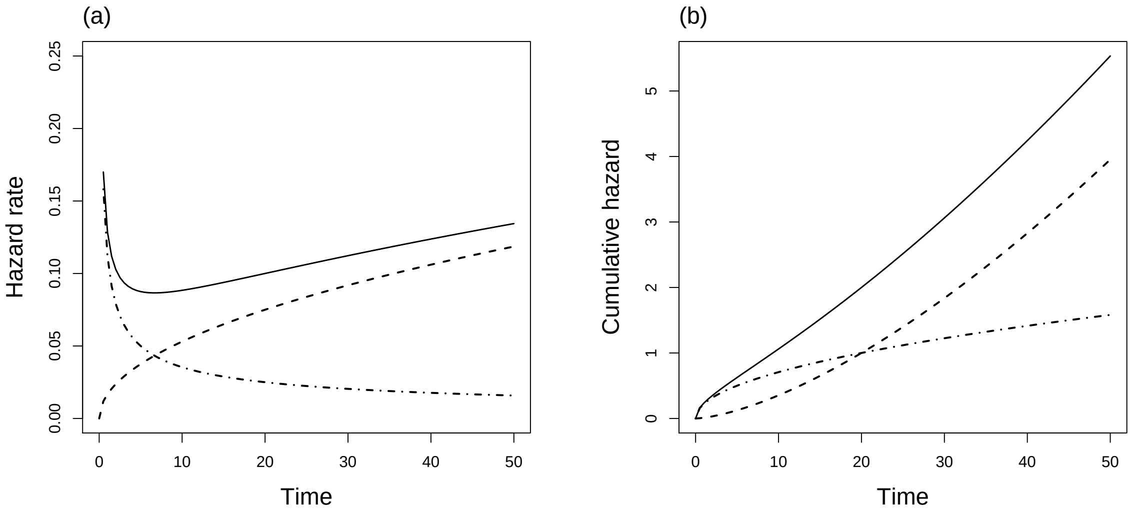

Figure 1: *(a) Hazard rate (—) and failure-specific hazard rates* $h_{0,1}$ *(- - -) and* $h_{0,2}$ *(-·-·-) in the prior. (b) The same for the cumulative hazard.*

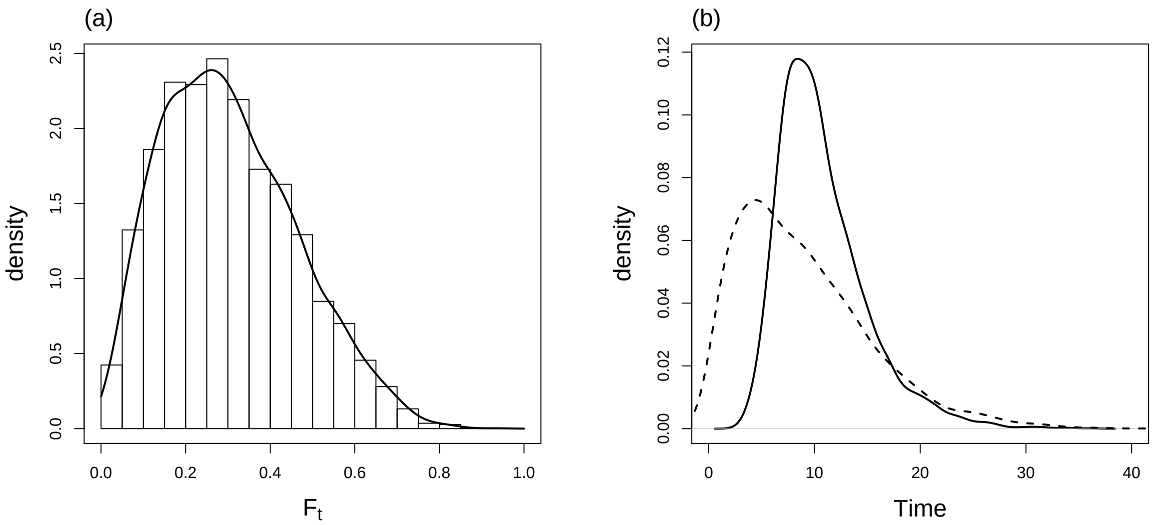

Figure 2: *(a) Histogram and density estimates of the posterior distribution of* F_t *for* $t = 5$*. (b) Density estimate of* $I_T(\tilde{\mu}^*)$ *(—) and* $I_T(\tilde{\mu})$ *(- - -) for* $T = 50$*; sample mean and sample s.d. of* $I_T(\tilde{\mu}^*)$